Reflections on Language Documentation
20 Years after Himmelmann 1998

edited by

Bradley McDonnell
Andrea L. Berez-Kroeker
Gary Holton

In search of island treasures: Language documentation in the Pacific

Alexandre François

LaTTiCe, CNRS–ENS–Sorbonne nouvelle
Australian National University

The Pacific region is home to about 1,500 languages, with a strong concentration of linguistic diversity in Melanesia. The turn towards documentary linguistics, initiated in the 1980s and theorized by N. Himmelmann, has encouraged linguists to prepare, archive and distribute large corpora of audio and video recordings in a broad array of Pacific languages, many of which are endangered. The strength of language documentation is to entail the mutual exchange of skills and knowledge between linguists and speaker communities. Their members can access archived resources, or create their own. Importantly, they can also appropriate the outcome of these documentary efforts to promote literacy within their school systems, and to consolidate or revitalize their heritage languages against the increasing pressure of dominant tongues. While providing an overview of the general progress made in the documentation of Pacific languages in the last twenty years, this paper also reports on my own experience with documenting and promoting languages in Island Melanesia since 1997.

1. Approaching language documentation in the Pacific   This paper reflects on twenty years of linguistic documentation in the Pacific. After an overview of the region's rich linguistic ecology (§1), I will survey the progress made so far in documenting and archiving valuable recordings from the region (§2). Crucial to the success of language documentation is also its relevance to speaker communities in their strive to preserve and revive their own languages (§3).

1.1 Overview of Pacific languages   The Pacific region is home to about 20% of the world's languages (Simons & Fennig 2018), and hosts a great number of different language

families. In terms of human and linguistic geography, the term *Pacific* commonly refers to the set of inhabited islands located within the Pacific Ocean, south of the 30° N parallel. Depending on the context, the term may also include the Philippines and Indonesia—usually considered part of SE Asia—as well as Australia; but these areas are covered by other chapters in this volume (see Arka & Sawaki (2018) for SE Asia; Singer (2018) for Australia in this volume). The present chapter will thus identify the Pacific (Figure 1) as the vast area defined by the three subregions of Melanesia, Micronesia and Polynesia.[1]
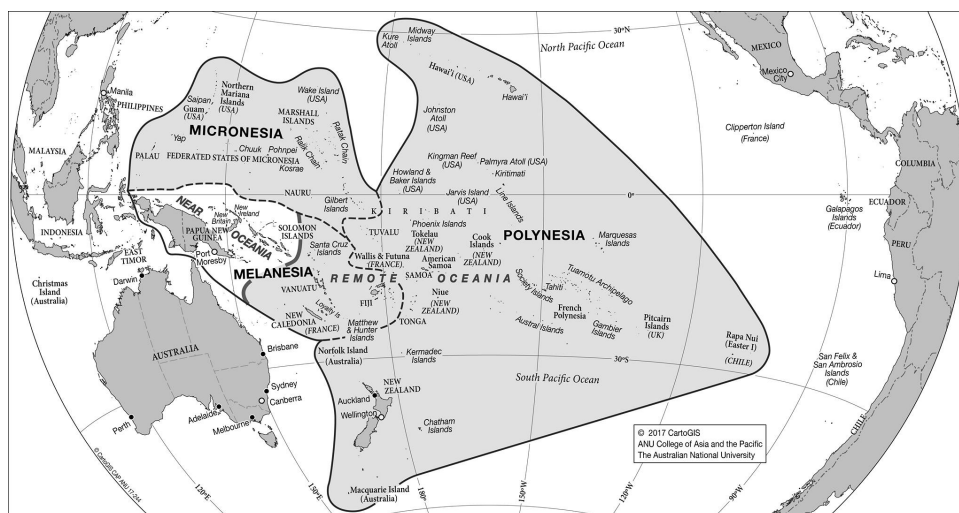

Figure 1: A map of the Pacific region, showing the subdivision into Melanesia, Micronesia, Polynesia; and the archaeological divide between Near Oceania (white) and Remote Oceania (gray) [© ANU College of Asia and the Pacific, CartoGIS, 2017]

The names for these three subregions were first introduced by Dumont d'Urville in 1831, and do not reflect accurately the populations' prehistory (Green 1991, Tcherkézoff 2009). More recently, Pawley & Green (1973) proposed to divide the same region based on archaeological criteria, into two areas:

- *Near Oceania* (consisting of the island of New Guinea and most of the Solomon Islands) was first settled by Homo sapiens more than 50,000 years ago;

- *Remote Oceania* (the rest of the Pacific—see grayed area in Figure 1) was only settled during the last 4,000 years.

As Map 1 shows, the boundary between Near and Remote Oceanic splits apart the area traditionally labelled as "Melanesia" (Green 1991).

The two areas also differ in their linguistic make-up. The more recently settled Remote Oceania features only one family, namely Austronesian—more exactly, the *Oceanic* branch of the Austronesian phylum. As for Near Oceania, it is home to about 80 genealogically unrelated language families and isolates,[2] making it the world's genealogically most

---

[1] Map provided by CartoGIS Services, ANU College of Asia and the Pacific, The Australian National University.
[2] "The Papuasphere […] contains, by the current count, 862 languages comprising 43 distinct families and 37 isolates." (Palmer 2018:6).

diverse area. These various families are collectively referred to using the umbrella term *Papuan*, to which one must add a more recent layer of Austronesian migrations.

Table 1 shows the unequal distribution of indigenous languages across the different subregions of the Pacific.[3] Micronesia and Polynesia, which were settled relatively recently, show lower linguistic diversity, with only about 60 languages together for such a vast area. By contrast, the region of Melanesia shows considerable diversity with a sheer total of 1419 languages—whether Austronesian (557) or non-Austronesian (862).

Area	Language family	# languages
MELANESIA, Near Oceania (New Guinea, Solomon Islands)	"Papuan" (≈80 families)	862
MELANESIA, Near Oceania (New Guinea, Solomon Islands)	Austronesian	345
MELANESIA, Remote Oceania (eastern Solomon Island, Vanuatu, New Caledonia, Fiji)	Austronesian (Oceanic)	212
POLYNESIA	=	38
MICRONESIA	=	21
Pacific region		**1478**

Table 1: Distribution of languages across the different subregions of the Pacific

1.2 Language density and vitality According to Simons & Fennig (2018),[4] Pacific languages, considered as a whole, show an average of 5,271 speakers per language, with a median value of 980. These figures are the lowest of all continents: they can be compared, respectively, with the world's average of 1 million speakers per language, with a median value of 7000.

One can in fact observe extreme discrepancies in language density between different areas of the Pacific (Pawley 1981, 2007). On one extreme, the language with most speakers is Sāmoan, with 413,000 speakers (*Ethnologue*). On the other extreme, the average size of a language community in Vanuatu at the beginning of the 20th century—at a time when the archipelago went through a demographic bottleneck—was "as low as 565 speakers per language" (François et al. 2015: 9). This goes to show the drastic gap in the language ecology across different parts of the Pacific—as "Melanesian diversity" (Dutton 1995, Unseth & Landweer 2012) contrasts so strongly with "Polynesian homogeneity" (Pawley 1981).

The diversity of Papuan languages may arguably be explained by the considerable time depth of human settlement in Near Oceania, and a long history of migrations and language evolution. But what is perhaps more striking is that even the Austronesian-speaking populations, which have had less than four millennia of *in situ* development, achieved

---

[3]Numbers are taken or inferred from the database Glottolog 3.2 (Hammarström et al. 2018). Usual disclaimers apply when counting languages. Also, note that "38" is the number of languages belonging to the Polynesian branch of Oceanic languages: only about half of these are spoken within the *Polynesian triangle* (the area labelled 'Polynesia' on Figure 1), while the remainder, known as *Polynesian outliers*, lie geographically in Micronesia or Melanesia.

[4]See https://www.ethnologue.com/statistics.

a similar rate of language density. Vanuatu, for example, was first settled 3,100 years ago by speakers of Proto Oceanic (Bedford & Spriggs 2008); and in that relatively short time-span, the archipelago's small population (currently 0.3 million) managed to diversify into 138 distinct languages—making it the country with the world's highest linguistic density *per capita* (François et al. 2015).

1.3 Different landscapes, different strategies This overview of the varying linguistic landscapes found across the Pacific (§1.1) entails quite different approaches when it comes to language documentation.

The languages with larger speaker populations numbering over 100,000, such as those found in major Polynesian centers (Sāmoan, Tongan, Māori, Tahitian, etc.) are certainly threatened in the long term due to the pressure of colonial languages—French, English—and of globalizing trends; but for the immediate future, they can be deemed safe from immediate endangerment. Because these languages have already been the object of grammatical or lexical descriptions, the work of linguists is rather to document the various styles and registers of these languages—whether that be technical vocabulary, verbal art, poetry (e.g. Meyer 2013 for Tahitian)—or the internal dynamics of their variation (Love 1979, Duranti 1997 for Sāmoan).

The situation is different with the many languages of Melanesia, or indeed with the demographically smaller languages of Micronesia or Polynesia. About half of Pacific languages are spoken by populations below the threshold of 1,000 speakers, which makes them more vulnerable to the risk of language shift and loss. In view of the sheer number and diversity of smaller languages of the Pacific, the urgent task is often for linguists to describe and document the linguistic practices of these speech communities while the languages are still vital. The last two decades have seen considerable effort in that direction; and while a lot remains to be done in the region, it is already possible to report on various successful endeavors in the domain of language documentation in the Pacific. The next sections will illustrate some of these efforts, and outline ways in which they can be appropriated by language communities.

2. Fieldwork archives

2.1 From description to documentation A few decades ago, the work of missionaries and pioneer linguists typically consisted in collecting basic data in the form of wordlists (e.g. Leenhardt 1946 for New Caledonia, Tryon 1976 for Vanuatu) or grammar sketches (e.g. Codrington 1885, Ray 1926 for Melanesian languages). Apart from translations of the Scriptures, it was rare to collect or publish texts, or other samples of connected speech.

The tide turned when linguists understood that their role was to record languages in the way they were actually spoken. Rather than eliciting wordlists or translating grammarians' sentences, language describers began conscientiously collecting high-quality data. This involves recording spontaneous speech in various forms: narratives, procedural texts and explanations, personal memories, conversations. That important evolution had already begun in the 1970s—as witnessed, for example, by the volumes of stories collected in various languages of Melanesia (e.g. Ozanne-Rivierre 1975–79; Paton 1979; Bensa & Rivierre 1982; Facey 1988) and Polynesia (e.g. Frimigacci et al. 1995). In the same spirit, researchers and engineers at Paris-based CNRS–LaCiTO created the first online

audio archive in endangered languages, as early as 1996 (Jacobson, Michailovsky & Lowe 2001), bringing together valuable fieldwork recordings with their text annotations.[5]

Himmelmann's explicit proposal for *language documentation* (Himmelmann 1998) thus came at a timely point in the evolution of researchers' practices. Rather than giving value to the sole results of linguistic analysis in the form of academic papers, grammars or dictionaries, the focus was shifting towards the quality and availability of actual samples of spontaneous speech in the various languages under study.

The rationale for the new focus on high-quality linguistic data was manifold. The insistence on gathering spontaneous speech serves an aesthetic criterion—the wish to pay tribute to the world's intangible linguistic heritage—but also a scientific one. If linguistic description and typology are meant to be an empirical science, it is not sufficient to translate pre-hashed questionnaire sentences derived from a linguist's theoretical model; instead, it is essential that the object of our scrutiny exists in the form of an observable corpus, independent of any aprioristic prejudice on what sort of structures we should expect to find.

2.2 Language archives Pacific languages are well represented in a number of online archives. The OLAC *Language Resource Catalog*[6] lists several collections featuring Pacific resources. Some of these archives, like *Rosetta* or *SIL-LCA*, include simple wordlists, grammar sketches, or translated sections of the Scriptures. Only some of the archives here mentioned pertain to documentary linguistics strictly speaking, in the sense of linguistic corpora based on naturalistic speech. The form usually taken by these language resources is as audio or video recordings, of varying length, typically ranging between 1 and 20 minutes each. These media are archived either as raw sound recordings, or as sound enriched with text annotations: transcription and free translation—with the additional possibility of interlinear glossing.

Table 2 lists the number of audio recordings[7] featured in the catalog of the world's four main language archives dealing with the Pacific: PARADISEC (Thieberger & Barwick 2012); DOBES and other databases hosted by the MPI (Brugman et al. 2002, Wittenburg et al. 2002); CNRS–LaCiTO's *Pangloss Collection*[8] (Jacobson et al. 2001; Michailovsky et al. 2014); and University of Hawai'i's *Kaipuleohone* (Albarillo & Thieberger 2009; Berez 2013). I only list resources in indigenous languages of the Pacific, including pidgins and creoles, to the exclusion of colonial languages.

Recordings from Pacific languages are more or less prominent in each archive. For example, the 3039 Pacific audio samples found in the MPI Language Archive correspond to a mere 5 percent of their entire catalog of recordings. By contrast, Pacific languages

---

[5]LaCiTO's online archive, which was later named the *Pangloss Collection* (Michailovsky et al. 2014), was initially developed by Boyd Michailovsky, Martine Mazaudon and John Lowe, and later expanded by Michel Jacobson. Pangloss is the largest collection within the CoCoON repository—see fn.8.

[6]OLAC Language Resource Catalog: `http://dla.library.upenn.edu/dla/olac/`. Note that ELAR, the *Endangered Languages Archive* developed at SOAS (Nathan 2010), is not featured under OLAC, and is thus unfortunately absent from the present statistics; yet that archive contains 36 archival deposits from the Pacific. Particularly noteworthy is Mike Franjieh's collection on Northern Ambrym languages (Franjieh 2018), which earned Delaman's Franz Boas 2019 award for best online multimedia documentary collection.

[7]This corresponds to the type 'Sound' among the categories proposed by the *Dublin Core Metadata Initiative* used by OLAC. Similar statistics could be carried out with video (DCMI 'MovingImage'), yet in this paper I will restrict myself to audio resources, for the sake of brevity and consistency.

[8] In the OLAC catalog, the *Pangloss Collection* appears under the name "CoCoON" (*Collections de Corpus Oraux Numériques*), a compilation of several audio archives hosted by CNRS' Huma-num infrastructure. LaCiTO's Pangloss is the largest collection within CoCoON, and the only one dealing with Pacific languages.

Archive	Institution	# Audio in catalog	# Audio from Pacific	% Catalog
PARADISEC	U Melbourne, ANU, U Sydney; CoEDL	9748	4458	46 %
DOBES, MPI-PL, Language collections	MPI for Psycholinguistics (Nijmegen)	66721	3039	5 %
Pangloss Collection	CNRS–LaCiTO	3302	1422	43 %
Kaipuleohone	U Hawaiʻi	2571	804	31 %

Table 2: Language archives displaying audio resources in indigenous Pacific languages

are proportionally better represented in PARADISEC (which has *Pacific* in its very name) and in the *Pangloss Collection*.

Archives differ in their precise geographical coverage. As Table 3 shows, languages of Papua New Guinea (whether Papuan or Austronesian) are well represented in PARADISEC and MPI-DOBES. Those of Vanuatu are mostly found in PARADISEC and Pangloss. Pangloss is also the place to go for languages of New Caledonia. Kaipuleohone's strong spots are PNG and Micronesia.

	PARADISEC	MPI-DOBES	Pangloss	Kaipuleohone	Total
Melanesia: Pidgins & creoles	147	25	13	2	*187*
PNG, Solomons: *Papuan*	1953	1806	–	126	*3885*
PNG: *Austronesian*	861	686	–	354	*1901*
Solomons: *Austronesian*	459	178	93	33	*763*
Vanuatu	874	113	869	–	*1856*
New Caledonia	27	–	404	3	*434*
Fiji	88	–	–	50	*138*
Micronesian	17	–	–	212	*229*
Polynesian	32	231	43	24	*330*
TOTAL	4458	3039	1422	804	**9723**

Table 3: Number of media resources for each archive, organized by geographic and linguistic area

2.3 A sample of individual corpora Table 4 lists the ten richest audio corpora—judging by the number of media resources—for individual languages of the Pacific. The four main archives (cf. Table 2) are represented, as well as the four main countries making up Melanesia.

Among these online corpora, those hosted by the MPI, Kaipuleohone or Paradisec, are difficult to access as they require special authorization, even for consultation. The most readily accessible resources are those of the *Pangloss Collection*, which makes it easier to

Language	Family	Country	Linguist	Archive	# audio
Yelî Dnye	(Papuan)	PNG	S. Levinson	MPI	722
Mwotlap	Oceanic	Vanuatu	A. François	Pangloss	504
Savosavo	(Papuan)	Solomons	C. Wegener	MPI	462
Saliba	Oceanic	PNG	A. Margetts	MPI	423
Bebeli	Oceanic	PNG	H. Sato	Kaipuleohone	356
Titan	Oceanic	PNG	T. Schwartz	PARADISEC	327
South West Bay	Oceanic	Vanuatu	L. Dimock	PARADISEC	302
Blablanga	Oceanic	Solomons	R. Voica	PARADISEC	292
Teop	Oceanic	PNG	U. Mosel	MPI	234
Cèmuhî	Oceanic	New Caledonia	JC Rivierre	Pangloss	231

Table 4: The ten richest audio corpora for individual languages from the Pacific (source: OLAC)

find out about statistics. I will give a brief overview of the Cèmuhî and Mwotlap corpora available there.

Cèmuhî, a tonal language of New Caledonia, is represented on Pangloss by 231 audio resources, recorded in the field between 1965 and 1979 by the late Jean-Claude Rivierre.[9] Each entry is a WAV file, digitized from legacy reel-to-reel tapes. Most recordings feature traditional narratives, whether myths or folktales (see Rivierre & Ozanne-Rivierre 1980, Bensa & Rivierre 1982). Among these 231 audio entries, 56 are accompanied by text annotations, consisting of a transcription, a free translation, and glosses—see Figure 2.

Mwotlap, a language of the Banks Islands (north Vanuatu), is featured in 504 audio recordings, which I collected between 1997 and 2011. With about 52 hours in total, the Mwotlap corpus forms about half of the recordings I have archived on Pangloss. Altogether, these consist of 962 resources, totalling 104 hours of sound,[10] in twenty-three different languages: four Oceanic languages of the Solomon Islands (Lovono, Tanema, Teanu, Tikopia), eighteen Oceanic languages of Vanuatu (Araki, Dorig, Hiw, Koro, Lakon, Lehali, Lemerig, Lo-Toga, Löyöp, Mota, Mwerlap, Mwesen, Mwotlap, Nume, Olrat, Vera'a, Volow, Vurës), and one creole (Bislama).

My archives on Pangloss take the form of sound files with metadata, downloadable under a Creative Commons licence (see fn.11). About 105 resources (including 38 for Mwotlap) are accompanied by time-aligned transcriptions and other text annotations, similar to Figure 2 above.

About one third of the archived resources are musical performances of various sorts, from dances to sung poetry: these formed the basis of a discographic publication together with the ethnomusicologist Monika Stern (François & Stern 2013). The remaining two thirds (69%) represent connected speech—mostly folk narratives (389 stories), but also procedural explanations, conversation, elicitation. A selection of narratives was the

---

[9]Besides Cèmuhî, Rivierre has also archived recordings in three other Kanak languages: Paicî, Numèè and Bwatoo—see `http://tiny.cc/Rivierre-archives`.
[10]Link: `http://tiny.cc/Francois-archives`.

Figure 2: Screenshot of the Cémuhî corpus: The traditional story "Les écailles de poisson de Tiwécaalè" is presented in a time-aligned transcription, with translation and glossing (Jean-Claude Rivierre, LaCiTO–CNRS; story by Bernadette Tyèn).

source of several booklets I created towards the consolidation of vernacular literacy in various speaker communities (§3.5).

3. Community outreach Language documentation not only involves the work of linguists, but also favors initiatives by speaker communities towards the preservation and revitalization of their heritage languages.

3.1 Enhancing access to the resources The audio or video resources made by linguists can be highly valued by the community of speakers. These documents preserve the memory of specific individuals, storytellers or singers or personalities who can now be remembered by their relatives, descendants and countrymen. The recordings also encapsulate cultural knowledge, folk traditions, oral history, important narratives and artistic forms that deserve to be passed on to the next generations. Finally, they also capture the various shapes taken by a living language, whether in the form of dialogues, stories, or verbal art; obviously, this linguistic testimony is all the more precious when the language is threatened with extinction within a few decades or years. For all these reasons, the current speakers of the language, or their descendants to come, constitute a key audience for our documentary work (Turin et al. 2013).

One aspect to be developed in the near future is the ease with which community members can access our archives, even when they are not technically savvy, or familiar with linguists' circles. Archives on a particular language should turn up in public search engine results, and at least the catalog of resources be easily and intuitively searchable. Most archives require the end user to create an account, and often to ask for permission to access specific resources: while this may be necessary in some cases, this is often a *de facto* obstacle to many community members who would like to casually listen to their heritage language without having to go through the intimidating process of an official request. In this sense, a fully public option certainly has its advantages.

LaCiTO's Pangloss Collection (or its sister CoCoON: see fn.8) is the only online archive that seems to fill those requirements at the moment: even if its interface could still be made more appealing to a lay audience, at least it is intuitive enough that anyone can easily navigate it, retrieve some resources, and listen to them right away—since they are all provided in Open access[11] and require no authorization.

One important advantage of this easy access is the possibility to share specific recordings with community members, e.g. via social media. In the last years 2014-2017, I created Facebook pages for the communities speaking respectively the languages of Araki, Mwotlap, Hiw (Vanuatu) and Teanu (Solomon Islands). Besides promoting the use of vernacular languages in writing (§3.5), each page also gives us the opportunity to share links to individual archived fieldwork recordings. Occasionally, I can send the link to a resource as a reply to a member's specific request—whether they're looking for a particular story, or an old song, or a sample of their grandfather's voice. The possibility to access recordings so readily was always warmly welcomed by group members. This is a simple and efficient way to return the fruit of our research to the younger members of the communities, in a way meaningful to them.

3.2 Searching across corpora and archives Documentary resources on Pacific languages are currently distributed across different archives, each with its own interface, principles, technical options. A certain level of inter-operability across archives has already been achieved through the common adoption of the standards of *Dublin Core Metadata Initiative*, and the shared connection with the OLAC initiative (Simons & Bird 2003). This has made it possible to cross-search several archives at once based on geographical or linguistic criteria, and navigate from one multimedia repository to another, across institutional boundaries.

However, at the moment this cross-integration among archives remains limited, and would benefit from being enhanced. Suppose an anthropologist, or a Pacific islander, would like to scan all existing archives from a certain region of the Pacific, say for stories containing the terms "canoe", or "shark", or "magic" (either in the title, or in the text's English translation). At the moment, many individual repositories lack a search function or a concordance tool, which is definitely a gap to fill. A community member, or a scholar, may wish to search one or several text corpora for a certain word form or gloss—a simple endeavor that is often still impossible. Such a search engine could also make it possible to look for an important placename, or to retrieve a valuable story which had ceased to be transmitted. The title proposed in the metadata—often the only searchable segment—is not always sufficient. Ideally, the tools for navigating or interrogating corpora would be pooled together (technology permitting) across different archives.

---

[11] The default license at Pangloss is CC BY-NC-ND 3.0 (Attribution, Non-commercial, No Derivatives). Recordings that are sensitive for cultural or social reasons are not displayed publicly on Pangloss.

3.3 Repatriating recordings Like in most places of the world (Pearce & Rice 2013), the internet in the Pacific is nowadays less and less accessed through computers, and more frequently through mobile interfaces; this may well inform our practices in terms of designing our tools in the future. Yet in spite of fast improvements, in many rural areas of the Pacific internet access still remains costly and unreliable, so much that offline solutions are still welcome for the diffusion of knowledge.

In 2011, I thus chose to repatriate all my field recordings to speaker communities in the form of a local digital copy—first, to the Vanuatu Cultural Centre in Port Vila; and secondly, to a newly created *Torres-Banks cultural centre* on the island of Motalava (François 2012; Michailovsky et al. 2014: 131). Vanuatu's Alliance Française helped maintain this cultural centre, funded a laptop, and provided technical training to local curators.

Admittedly, searching through a thousand audio recordings in 23 languages, from many locations, with so many genres and contributors, would constitute a challenge for the local users, most of whom had never used a computer. For that reason, I designed an intuitive way to search through the archives. I exploited the possibilities offered by free media players such as iTunes or Winamp, then available on local interfaces. Each archived recording was converted from wav into an mp3 file. The latter was then enriched with id3 metadata, which were automatically imported from Dublin Core describers as already stored in the Pangloss archive: e.g. the name of the speaker or storyteller became the "ARTIST", the place of recording was mapped onto the "GROUPING" tag, and so on. Each recording was associated with a photograph—generally, a portrait of the speaker—which took the place of an album's artwork. The result was an enticing multimedia digital library featuring hundreds of recordings, that could be explored either through pictures, or through a multifaceted search involving: *language; title; name of speaker or musician; date of recording; location; genre; duration* (François 2012).[12] This local search interface was welcomed by the community members, who were able to search through the collection in many ways, and retrieve recordings of their interest on the computer—whether samples of speech or music. They would then download these recordings from the public media library to their mobile phones used as offline mp3 devices, and distribute them to their friends and families. This proved a successful way to share digital resources amongst the community, even in an offline context.

Back in 2011, I found it technically difficult to replicate the same rich search interface using online tools. However, given the current spread of the internet, and the latest developments of mobile applications, it should now become easier to come up with user-friendly search tools so as to distribute the fruit of our documentary work to non-academic users through mobile-based devices.

3.4 Community-driven documentation We just saw how communities can benefit from the efforts of language documentation carried out by linguists, through increased access to valuable recordings. Interestingly, these results can also inspire the speakers themselves to pursue the work of documentation on their own language.

While the description of a language's grammar or lexicon requires solid academic training in the domain of linguistics, the documentation of linguistic practices is a different sort of endeavour, whose crucial ingredients include: familiarity with the language to be documented, and ability to transcribe it; acquaintance with the cultural universe attached

---

[12]The video at `https://youtu.be/hZGm0CLzxU8` demonstrates the potential of that interface.

to it; understanding of what is at stake in language endangerment and documentation; personal motivation. When these conditions are present, native speakers are often in a good position to bring about language documentation themselves. They can then set out to record the language under its various manifestations—whether traditional narratives, procedural texts, conversations—using audio or video technologies (Carew et al. 2015, Bettinson & Bird 2017).

Recent technological developments have made access easier to decent-quality microphones, including through the use of commercial smartphones. Some apps have been developed for the creation of dictionaries—such as *Ma! Iwaidja* for the Iwaidja language of northern Australia, or *Ma! BenaBena*, with respect to a Papuan language of Papua New Guinea (Birch 2013). Their interface allows crowdsourcing, and the enrichment of lexical data by the community members themselves (Carew et al. 2015: 314).

Another tool tailored for community-driven documentation is the Android app "Aikuma" (Bird et al. 2013; Bettinson & Bird 2017), itself superseded by Lɪɢ-*Aikuma* (Blachon et al. 2016). This user-friendly application allows native speakers to enrich an existing recording, or a newly-created one, using vocal annotation, such as slow-speech "respeaking" or translation. The interface is designed to be used intuitively by community members even if they are not literate. The software was first used with speakers of Usarufa in Papua New Guinea, a language where *aikuma* means "meeting place".[13].

More recently, *Aikuma* has also become the name of a collaborative project, with the aim to promote the celebration of indigenous languages through oral performances of storytelling and verbal art. The project has been active online: `https://twitter.com/AikumaProject`.

3.5 Language learning and revitalization   Among the many positive outcomes of language documentation projects, is their possible usefulness for language revitalization and language learning.

Through collaborative workshops and initiatives, web-based projects and team efforts, an increasing number of activities are taking place across the Pacific—like elsewhere in the world—that promote the use of vernacular languages in speaking and in writing. One could cite the cases of Māori *kōhanga reo* or "language nests" in New Zealand (Benton 1989, King 2001); of *pūnana leo* for Hawaiian (Warner 2001); or similar attempts in French territories, whether French Polynesia (Paia 2014) or New Caledonia (Moyse-Faurie 2012, Vernaudon 2015)… These revitalization projects are only tangentially related to documentary linguistics per se (see Penfield & Tucker 2011), but they participate in a general push to increase the exposure of younger generations to their legacy languages in the variety of their manifestations. In some cases, efforts in language learning and revitalization are directly linked to the enterprise of linguistic documentation, and with the recording of spontaneous speech from fluent speakers.

Indeed, the high-quality samples of fluent speech, such as the audio and video resources found in archives, deserve to be exploited for their teaching potential. The target audience may be outsiders wishing to learn a new language; or members of the language community—whether speakers or semi-speakers themselves—wishing to access valuable recordings in their heritage languages. Complete beginners would first need to have access to teaching materials; but once they've acquired enough fluency to understand

---

[13]See `http://www.aikuma.org/faq.html`

simple stories, then the audio or video recordings—especially if enriched with annotations and glosses—are of considerable help to increase their linguistic competence.

Accessing the audio or video recordings has a great learning potential, especially for learners who wish to maintain fluency in their heritage language, by hearing high-quality storytelling from their elders. But another way in which language documentation can be used in a teaching context, is by developing skills in literacy, whether in regular schools or in community-led learning groups.

While many possible examples could be cited, I will briefly report on my personal experience in Melanesia. The majority of the 23 languages on which I carried out language documentation in Vanuatu and the Solomons (§2.3) lacked any stable orthography when I began working on them. One of my roles, as a linguist, has been to sort out the phonology of each language (François 2011a: 194), and design a set of spelling conventions, which I then discuss with community members during public meetings. Once a system has been agreed upon, I can start sharing the transcriptions of my field recordings with speakers. In many cases, this exchange proves a pivotal moment for the community, whose oral language is finally endowed with an "official" orthography.

The next stage is to create teaching materials to make sure community members can master the spelling system that was agreed upon. The effort required for this learning depends a lot on the difficulty of the language's phonology—particularly, its vowel inventory. Because Teanu, the main language of Vanikoro (François 2009), has only five vowels /i e a o u/—the same as those in the Roman script—its speakers master the orthography very fast. But it takes more time, and more exposure to written materials, in order to transcribe consistently such languages as Hiw with its nine vowels /i ɪ e ɵ ə a ʉ o ɔ/; or Lemerig with eleven /i ɪ ɛ æ a œ ø ɒ ɔ ʊ u/.

Over the years, I have produced a number of booklets for literacy education. All volumes are also available on my homepage in digital format (François 2015)—a total of 21 books so far, of two different kinds. Ten books took the form of a basic alphabet primer, exemplifying each grapheme with a selection of words and phrases, with rich homemade illustrations. The remaining eleven volumes are story books, and constitute readers for a more advanced level of literacy. Their list is provided in Table 5.

Apart from one reader whose text I wrote in Mwotlap with the help of Edgar Howard (François & Howard 2000), all volumes showcase a selection of traditional narratives in each language, taken from my corpus of transcribed stories. Their length ranges from 36 to 78 pages each. They are monolingual, as they aim to encourage literacy in the vernacular language, as opposed to the country's dominant languages of education (French, English or Bislama) which are often people's default choices when it comes to writing. Figures 3 and 4 illustrate one page from two of those story books.

The literacy materials in question were initially self-published and self-funded by my family. In 2006, the Vanuatu bureau of Alliance Française added their support to six of these volumes, by funding the printing of 300 copies, as well as their shipping to different local schools in the Banks group; we renewed our collaboration again in 2015, with respect to four books destined to the communities of the Torres islands. Since then, I have heard numerous reports that these literacy materials have been successfully perused in various schools of the Torres-Banks province of Vanuatu—whether this was an initiative of the local teacher, or an application of the country's recent pledges to develop vernacular literacy programs in early school years (Vanuatu Ministry of Education 2012).

Language	Area	Year	Book title	Contents
Araki	Espiritu Santo	2008	*Sorosoro m̃ar̄an Raki*	Stories in Araki (south Santo)
Dorig	Gaua, Banks	2011	*O susrig ble mraw vata Dōrig*	Stories in Dorig
Hiw	Torres Is.	2015	*Ne vegevage r̄ossë 'n Mer̄avtit*	The legend of the hero Megravtit
Lemerig	Vanua Lava, Banks	2006	*Nvāv ʻām ʻa Lēmērig*	Stories in Lemerig
Lakon	Gaua, Banks	2011	*Suusuu pule maraw avan Lakon*	Stories in Lakon
Lo-Toga	Torres Is.	2015	*Ne vegevage së te Lō mi ne Toge*	Stories from Lo and Toga islands
Mwesen	Vanua Lava, Banks	2004	*O olñevu ta turm̃o ta M̄ēsēn*	Stories in Mwesen
Mwotlap	Banks	2003	*Tog tog i van en*	'Once upon a time': Stories in Mwotlap
Mwotlap	Banks	2000	*Bulsal, dam galsi me lēklek*	'Follow me, my friend': Language reader
Olrat	Gaua, Banks	2011	*Ususraa pule maraw men Ōlrat*	Stories in Olrat
Teanu +	Temotu, Solomon Is	2012	*Liatevo iepiene ñe piene akapa*	Stories in the three languages of Vanikoro: Teanu, Lovono, Tanema

Table 5: Text materials produced by the author based on his documentary work, aimed at literacy development and language revitalization

Now that the internet has reached speaker communities (since 2010 in the capital, since 2017 in some rural areas), I have witnessed increased use of vernacular languages in writing—whether in mailing lists, in texting, or on Facebook (§3.1). For some languages that used to show inconsistent attempts at transcription, it appears that younger contributors now show more consistency in their spelling. Some explained to me how they learned to write their own language using the literacy materials I produced, which gave them more confidence when writing their own legacy language.

52

Ba kē nemyōs so kē mas wuh kōyō, namyōs nonon so kē so niwuh kōyō.

Kē magal kōyō van so kōyō vanvan nen e tō kōyō mitiy, wa tō kē nivan hōw, tō wuh matmat kōyō mi naqyēñ.

Tō— kē nihatig hag nen tō— mahag van; kē net van so "Ohoo! Yoge gōh et mitimtiy te!" Kē menyē van, menyē van, tateh.

Ṁalṁal nonon Wētamat en nihohole van hiy nōlōmgep e wo "Imam mino tuwuh qiyig dōyō. Ba nēk valag hiy yow aslil en, nēk et van nibin lol tō anen, ba nēk he me nibin en, ba lep me noel."

Kē nilep nibin tiwag mi noel nen, van me nimōk veteg kōyō nen, vēsñēt kōyō van qele nōk, kōyō hohole!

"Hatig tō, dō van!" Mōk veteg nibin tiwag mi noel tō kōyō hohole noyō tō nen.

Mahē nognog meyen me ēgēn. Kōyō a hatig nen tō, lep namtehal tō a— van.

Kōyō van van van van va— van, van dēñ me lemtehal a nagayga dam tō aē a kēy mepyoy tēqēl tō kē aē en.
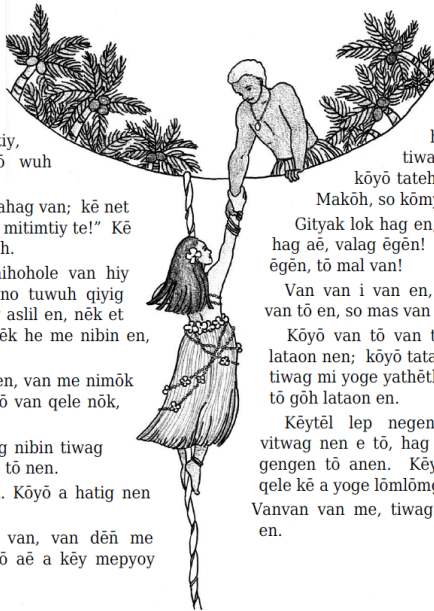
Wētamat yoñteg a so "Ooh! Vētmahē kē nimyen ēagōh!", so kē nivan wuh vege kōyō a kōyō nēh en. Kē nivan hōw me qele kē, nibin tiwag mi noel hohole tō nen, kōyō tateh. "Aah! Kōmyō so gal no? Makōh, so kōmyō ave?!"

Gityak lok hag en, lep nohos nonon, yow kal hag aē, valag ēgēn! Yoge mal dam ketket alge ēgēn, tō mal van!

Van van i van en, nēdēmdēm noyō so kōyō van tō en, so mas van dēñ a lataon.

Kōyō van tō van tō va— van, van dēñ hōw lataon nen; kōyō tatal kēlkēl laptō, imam nonon tiwag mi yoge yathēthēn en, mal van me, tō hag tō gōh lataon en.

Kēytēl lep negengen nagaytēl den nisto vitwag nen e tō, hag tō aslil lefranda nan nen, gengen tō anen. Kēytēl gengen en, so et van qele kē a yoge lōmlōmgep en nivanvan van me! Vanvan van me, tiwag mi ṁalṁal non Wētamat en.

Figure 3: A page from a story book in Mwotlap, Motalava I., Vanuatu (François 2003)
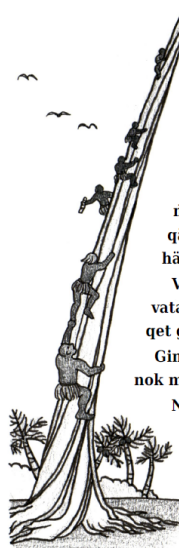
Wōrō ēn hag va— van, naw tē 'ät avōh, wōrō tē räkäs.

Räkäs mongēē wäl- wäl neñ, Qat gēn tē mäkä to sala nē ga to' jen nē.

Nē 'n häläqäg tägäh ṁērē heg, Tañro tē hälä- qäg kēl talvōn um, nē tē häläqäg talvōn.

Va—van, we tē wähä kere vata nagē Tañro neñ mēn too qet ga tutun we nē tē ginteg.

Ginteg hōw neñ, tutuan pah nok makē raṁos.

Nē tē row vēgēn hag, nē gēn tē luwluw gēn ek, sa *"Mok raṁos! Na ga marēs nēk a rēg! Nēk a vutgi!"*

We raṁos neñ tē rēg, tē vutgi ajew.

Gēē tē vēgēn suu-

suu. Tañro tē vēgēn suusuu gēē ma.

Gēē 'n vēgēn vēgēn va—n, Qat sa: *"Mok raṁos! Nēk rēg, nēk vutgi lēh hag!"*

Raṁos neñtē vutgi sa tē vutgi kakarañkē vata hag, hag neñ to lē pōlō.

Gēē 'n jēn hag neñ, sa *"Ha'ēh! Mok raṁos! Nēk ēt kaōl!"*

We 'n raṁos neñ tē kaōl.

Kaōl kaōl kaōl va—van, jēn hōw lē vanō neñ, nē tē häläqäg jen, tutuan tē häläqäg, nē tē gih lekteg uhli raṁos neñ.

Tañro neñ nē 'n kakal ma, va rigtäg we nē tē ukäg raṁos, raṁos tē 'äqä wutä i Qasval.

Figure 4: A page from a story book in Lakon – Gaua I., Vanuatu (François 2011b)

4. Conclusion: A mutual benefit for linguists and communities   The movement of documentary linguistics, as it emerged in the 1990s and was theorized by Himmelmann (1998), has meant a leap forward in the quality standards of the primary data serving as the basis for language description and analysis. In the Pacific region, this progress has endowed numerous endangered languages with rich corpora of naturalistic speech, in the form of audio and video recordings, with the frequent addition of text annotations. Future years should increase the mutual integration of language documentation and description, so as to reinforce the accountability of linguistic analyses based on solid empirical data.

Yet the relevance of language documentation goes beyond providing empirical material for the linguist. By focusing on naturalistic speech in different social contexts, the archives produced by the documentary enterprise have major potential for a broad array of audiences—anthropologists and historians, experts of oral literature and ethnomusicologists, language teachers and learners. Most importantly, the endeavor of documentary linguistics is proving of high relevance to members of Pacific speaker communities, whether their wish is to hear the voices of their elders, to create and peruse literacy materials in their legacy languages, or to hand over their linguistic and cultural knowledge to the upcoming generations.

References

Albarillo, Emily E., & Nick Thieberger. 2009. Kaipuleohone, University of Hawaiʻi's Digital Ethnographic Archive. *Language Documentation & Conservation* 3. 1–14.

Bedford, Stuart & Matthew Spriggs. 2008. Northern Vanuatu as a Pacific crossroads: The archaeology of discovery, interaction, and the emergence of the "ethnographic present." *Asian Perspectives* 47. 95–120.

Bensa, Alban, & Jean-Claude Rivierre. 1982. *Les Chemins de l'Alliance. L'organisation sociale et ses représentations en Nouvelle-Calédonie.* (Langues et Cultures du Pacifique.) Paris: Société d'Etudes Linguistiques et Anthropologiques de France.

Benton, Nena. 1989. Education, language decline and language revitalisation: The case of Maori in New Zealand. *Language and Education* 3(2). 65–82.

Berez, Andrea. 2013. The Digital Archiving of Endangered Language Oral Traditions: *Kaipuleohone* at the University of Hawaiʻi and *Cʼekʼaedi Hwnax* in Alaska. *Oral Tradition* 28(2). 261–270.

Bettinson, Mat, & Steven Bird. 2017. Developing a suite of mobile applications for collaborative language documentation. *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 156–164. Honolulu, Hawaiʻi, USA, March 6–7, 2017.

Birch, Bruce. 2013. The Ma! Project: Crowdsourcing software for language documentation. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of the Pacific and Regional Archive for Digital Sources in Endangered Cultures* `http://hdl.handle.net/2123/9858`

Bird, Steven, Florian Hanke & Haejoong Lee. 2013. Collaborative language documentation with networked smartphones. In Amanda Harris, Nick Thieberger & Linda Barwick (eds.), *Research, records and responsibility: Ten years of the Pacific and Regional Archive for Digital Sources in Endangered Cultures* `http://hdl.handle.net/2123/9857`

Blachon, David, Elodie Gauthier, Laurent Besacier, Guy-Noël Kouarata, Martine Adda-Decker & Annie Rialland. 2016. Parallel speech collection for under-resourced language studies using the Lɪɢ-Aikuma mobile device app. *Workshop on Spoken Language Technologies for Under-resourced Languages* (SLTU), May 2016, Yogyakarta, Indonesia. Procedia computer science.

Brugman, Hennie, Stephen Levinson, Romuald Skiba & Peter Wittenburg. 2002. The DOBES Archive: Its Purpose and Implementation. In Peter K. Austin, Helen Dry & Peter Wittenburg (eds.), *Proceedings of the international LREC workshop on resources and tools in field linguistics.* Paris: European Language Resources Association (ELRA).

Proceedings of the international LREC workshop on resources and tools in field linguistics. Paris: European Language Resources Association (ELRA).

Carew, Margaret, Jennifer Green, Inge Kral, Rachel Nordlinger & Ruth Singer. 2015. Getting in touch: Language and digital inclusion in Australian indigenous communities. *Language Documentation & Conservation* 9. 307–23.

Codrington, Robert H. 1885. *The Melanesian languages.* Oxford: Clarendon Press.

Duranti, Alessandro. 1997. Indexical speech across Sāmoan communities. *American Anthropologist* 99: 342–354.

Dutton, Thomas Edward. 1995. Language contact and change in Melanesia. In Peter S. Bellwood, James J. Fox & Darrell Tryon (eds.), *The Austronesians: Historical and comparative perspectives*, 207-228. Canberra: Australian National University.

Facey, Ellen E. 1988. *Nguna voices: text and culture from central Vanuatu.* Calgary: University of Calgary.

François, Alexandre. 2003. *Tog tog i van en* ['Once upon a time…']. Collection of stories from the oral tradition of Mwotlap. Monolingual in Mwotlap, for local use. Illustrated by Sawako François. Self-published. 78 pp.

François, Alexandre. 2009. The languages of Vanikoro: Three lexicons and one grammar. In Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross* (Pacific Linguistics 605), 103-126. Canberra: Australian National University.

François, Alexandre. 2011a. Social ecology and language history in the northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics* 1(2), 175-246.

François, Alexandre. 2011b. *Suusuu pule maraw avan Lakon* ['Traditional stories from Lakon']. Collection of stories from the oral tradition of Lakon (Gaua, Banks Is). Monolingual in Lakon, for educational use. Illustrated by Sawako François. Alliance Française de Port-Vila. 44 pp.

François, Alexandre. 2012. The media library project: Repatriating my audio archives. [Blog entry.] (`http://alex.francois.online.fr/AF-audio-library-e.htm`) (Accessed 2018-10-09).

François, Alexandre. 2015. Literacy materials for the communities: Building up vernacular literacy. [Blog entry.] (`http://alex.francois.online.fr/AF-literacy-e.htm`) (Accessed 2018-10-09).

François, Alexandre, Michael Franjieh; Sébastien Lacrampe & Stefan Schnell. 2015. The exceptional linguistic density of Vanuatu. In Alexandre François, Sébastien Lacrampe, Michael Franjieh & Stefan Schnell (eds.), *The Languages of Vanuatu: Unity and Diversity.* Studies in the Languages of Island Melanesia 5, 1–21. Canberra: Asia Pacific Linguistics Open Access.

François, Alexandre & Edgar Howard. 2000. *Bulsal, dam galsi me lēklek* ['Follow me, my friend']. Language reader for vernacular education, monolingual in Mwotlap. Illustrated by Sawako François. 26 pp. Reprinted in 2006 with the help of 'Alliance Française' of Vanuatu.

François, Alexandre & Monika Stern. 2013. *Musiques du Vanuatu: Fêtes et Mystères – Music of Vanuatu: Celebrations and Mysteries.* [Musical recordings] (Label INÉDIT). Paris: Maison des Cultures du Monde. CD album: 73'39", with booklet (128 pp.).

Franjieh, Michael. 2018. *The languages of northern Ambrym, Vanuatu.* Archive of linguistic and cultural material from the North Ambrym and Fanbyak languages. London: SOAS, Endangered Languages Archive. (`https://elar.soas.ac.uk/Collection/MPI1143013`) (Accessed 2018-10-20).

Frimigacci, Daniel, Muni Keletaona, Claire Moyse-Faurie & Bernard Vienne. 1995. *La tortue au dos moussue – Ko le fonu tu'a limulimua. Texte de tradition orale de Futuna.* Société d'Études Linguistiques et Anthropologiques de France 356. Louvain: Peeters.

Green, Roger. 1991. Near and Remote Oceania: Disestablishing "Melanesia" in culture history. In Andrew Pawley (ed.), *Man and a half: Essays in Pacific Anthropology and Ethnobotany in honour of Ralph Bulmer*, 491-502. Auckland: The Polynesian Society.

Hammarström, Harald, Sebastian Bank, Robert Forkel, & Martin Haspelmath. 2018. *Glottolog* 3.2. Jena: Max Planck Institute for the Science of Human History. (`http://glottolog.org`) (Accessed on 2018-10-18)

Harris, Amanda; Nick Thieberger & Linda Barwick (eds.), *Research, records, and responsibility: Ten years of the Pacific and Regional Archive for Digital Sources in Endangered Cultures.* Sydney: University of Sydney Press. `http://hdl.handle.net/2123/13310`

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36(1). 161–195.

Hinton, Leanne, & Kenneth Hale (eds). 2001. *The Green Book of Language Revitalization in Practice.* San Diego, CA: Academic Press.

Jacobson, Michel, Boyd Michailovsky, & John B. Lowe. 2001. Linguistic documents synchronizing sound and text. *Speech Communication* 33.1-2 (2001): 79-96.

King, Jeanette. 2001. Te Kōhanga Reo: Māori Language Revitalization. In Leanne Hinton & Ken Hale (eds.), *The Green Book of Language Revitalization in Practice* 118–128. San Diego, CA: Academic Press.

Leenhardt, Maurice. 1946. *Langues et dialectes de l'Austro-Mélanésie.* Vol. 46. Paris: Instut d'ethnologie.

Love, Jacob. 1979. Sāmoan variations. Cambridge, MA: Harvard University dissertation.

Meyer, David. 2013. *Early Tahitian poetics.* (Pacific Linguistics 641.) Berlin: Walter de Gruyter.

Michailovsky, Boyd, Martine Mazaudon, Alexis Michaud, Séverine Guillaume, Alexandre François & Evangelia Adamou. 2014. Documenting and researching endangered languages: The Pangloss Collection. *Language Documentation & Conservation* 8. 119-135. `http://hdl.handle.net/10125/4621`

Moyse-Faurie, Claire. 2012. Haméa et xârâgurè, langues kanak en danger. *UniverSOS: Revista de Lenguas Indígenas y Universos Culturales*, 73–86. Universitat de València.

Nathan, David. 2010. Archives 2.0 for endangered languages: From disk space to MySpace. *International Journal of Humanities and Arts Computing* 4(1-2). 111-124.

Ozanne-Rivierre, Françoise. 1975-79. *Textes nemi, Nouvelle-Calédonie.* 2 vol. Paris: SELAF.

Païa, Mirose. 2014. L'Enseignement des langues et de la culture polynésiennes à l'école primaire en Polynésie française. In Isabelle Nocus, Jacques Vernaudon, & Mirose Paia (eds), *L'École plurilingue en Outre-mer: Apprendre plusieurs langues, plusieurs langues pour apprendre*, 409–429. Rennes: Presses Universitaires de Rennes.

Palmer, Bill. 2018. Language families of the New Guinea area. In Bill Palmer (ed.), *The languages and linguistics of the New Guinea area: A comprehensive guide* (The World of Linguistics), 1–19. Berlin: de Gruyter.

Paton, W.F. 1979. *Customs of Ambrym (texts, songs, games and drawings)* (Pacific Linguistics D-22). Canberra: Australian National University.

Pawley, Andrew. 1981. Melanesian diversity and Polynesian homogeneity: a unified explanation for language. In Jim Hollyman & Andrew Pawley (eds.), *Studies in Pacific languages and cultures in honour of Bruce Biggs*, 259-310. Auckland: Linguistic Society of New Zealand.

Pawley, Andrew. 2007. Why do Polynesian island groups have one language and Melanesian island groups have many? Patterns of interaction and diversification in the Austronesian colonization of Remote Oceania. Paper presented at the conference *"Migrations"*, September 5–7, 2007, Porquerolles, Var, France.

Pawley, Andrew & Roger Green. 1973. Dating the dispersal of the Oceanic languages. *Oceanic Linguistics* 12(1/2): 1–67.

Pearce, Katy E., & Ronald E. Rice. 2013. Digital divides from access to activities: Comparing mobile and personal computer Internet users. *Journal of Communication* 63(4). 721–744.

Penfield, Susan D. & Benjamin V. Tucker. 2011. From documenting to revitalizing an endangered language: Where do applied linguists fit? *Language and Education* 25(4). 291–305.

Ray, Sidney H. 1926. *A comparative study of the Melanesian island languages.* Cambridge: Cambridge University Press.

Rivierre, Jean Claude, & Françoise Ozanne-Rivierre. 1980. *Mythes et contes de la Grande-Terre et des îles Loyauté (Nouvelle-Calédonie).* Paris: SELAF.

Simons, Gary & Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the world*, 21st edn. Dallas, Texas: SIL International. Online version: `http://www.ethnologue.com` (Accessed on 2018-10-18)

Simons, Gary & Steven Bird. 2003. The Open Language Archives Community: An infrastructure for distributed archiving of language resources. *Literary and Linguistic Computing* 18(2). 117–128.

Tcherkézoff, Serge. 2009. *Polynésie/Mélanésie: L'invention française des "races" et des régions de l'Océanie (xvie-xixe siècles).* Papeete: Au vent des îles.

Thieberger, Nick & Linda Barwick. 2012. Keeping records of language diversity in Melanesia: The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). In Nicholas Evans & Marian Klamer (eds.), *Melanesian languages on the edge of Asia: Challenges for the 21st Century*, 239–253. *Language Documentation & Conservation* Special Publication 5. `http://hdl.handle.net/10125/4567`

Tryon, Darrell T. 1976. *New Hebrides languages: An internal classification* (Pacific Linguistics C-50). Canberra: Australian National University.

Turin, Mark, Claire Wheeler & Eleanor Wilkinson (eds.). 2013. *Oral literature in the digital age: Archiving orality and connecting with communities.* Cambridge: Open Book Publishers.

Unseth, Peter & Lynn Landweer (eds.). 2012. *Language Use in Melanesia.* Special issue of *International Journal of the Sociology of Language* 214.

Vanuatu Ministry of Education. 2012. *Vanuatu National Language Policy.* Official report, 20 pp.

Vernaudon, Jacques. 2015. Linguistic Ideologies: Teaching Oceanic Languages in French Polynesia and New Caledonia. *The Contemporary Pacific* 27(2): 433–462. doi: 10.1353/cp.2015.0048

Warner, Sam L. Noʻeau. 2001. The movement to revitalize Hawaiian language and culture. In Leanne Hinton & Ken Hale (eds.), *The Green Book of Language Revitalization in Practice* 133–144. San Diego, CA: Academic Press.

Wittenburg, Peter, Ulrike Mosel & Arienne Dwyer. 2002. Methods of language documentation in the DOBES project. In Peter K. Austin, Helen Dry & Peter Wittenburg (eds.), *Proceedings of the international LREC workshop on resources and tools in field linguistics*, 36–42. Paris: European Language Resources Association (ELRA).

Alexandre François
alexandre.francois@ens.fr
ⓘ orcid.org/0000-0003-1947-0806